

CREDIT EDA CASE STUDY

By

Radhamani & Pranit



Problem Statement - Business Understanding

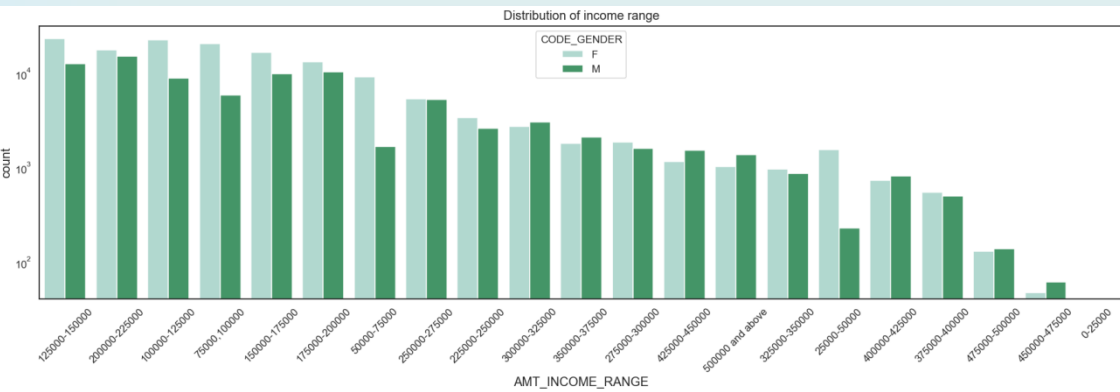
When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

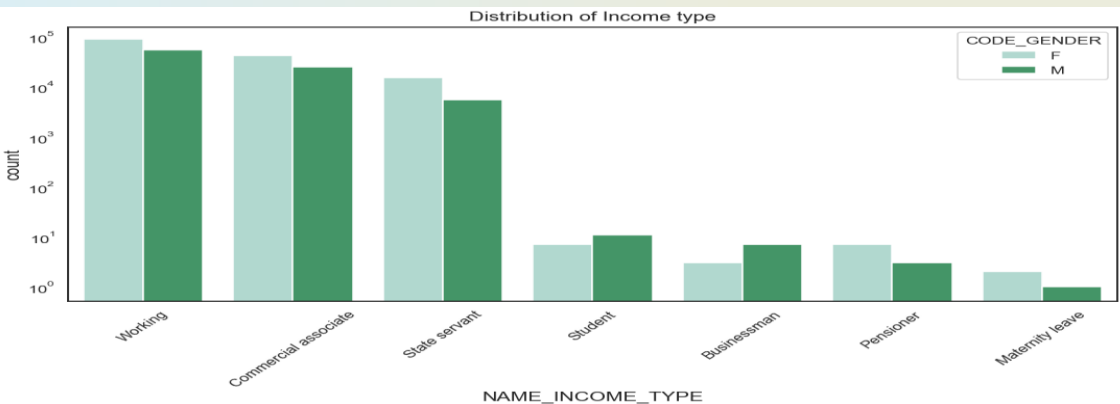
Approaching to Problem Statement

1. Reading & inspecting the Application and Previous application data
2. Cleaning the data by identify the missing values and using appropriate method to replace or remove
3. Identify if there are outliers in the dataset and their data imbalance.
4. Analyzing the imbalance using various graphs and explain the results of univariate, segmented univariate, bivariate analysis, etc. in business terms.
5. Conclusion

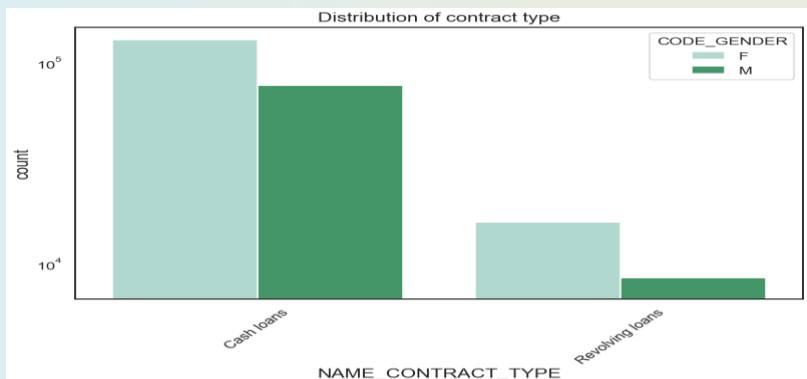
Univariate analysis for categories – Target0 (defaulters)



1. Females counts have high income range, specially more for the bin of 25000 to 70000.
2. Income range from 100000 to 200000 is having more number of credits.

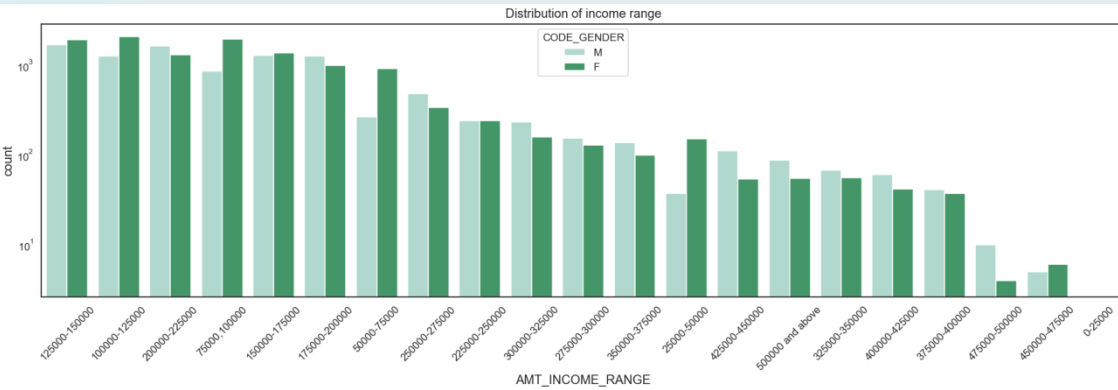


1. Working, commercial associate and State servant occupation type are having high income.
2. Here also females are high in count in most of the occupation

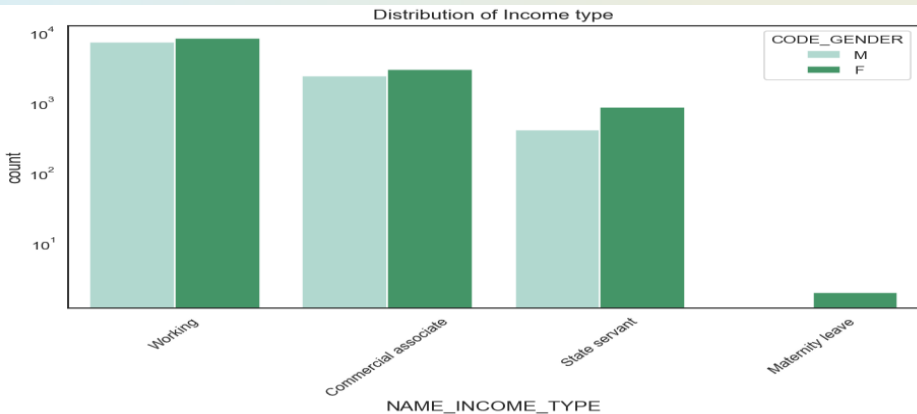


1. For contract type 'cash loans' is having higher than 'Revolving loans' contract type.
2. For this also Female is leading in applying loans.

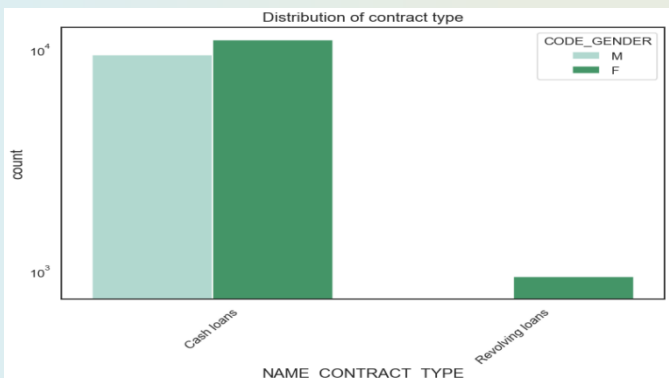
Univariate analysis for categories – Target1 (non- defaulters)



1. Female counts are slightly higher side of income than the Male.
2. Income range from 75000 to 225000 is having more number of credits.

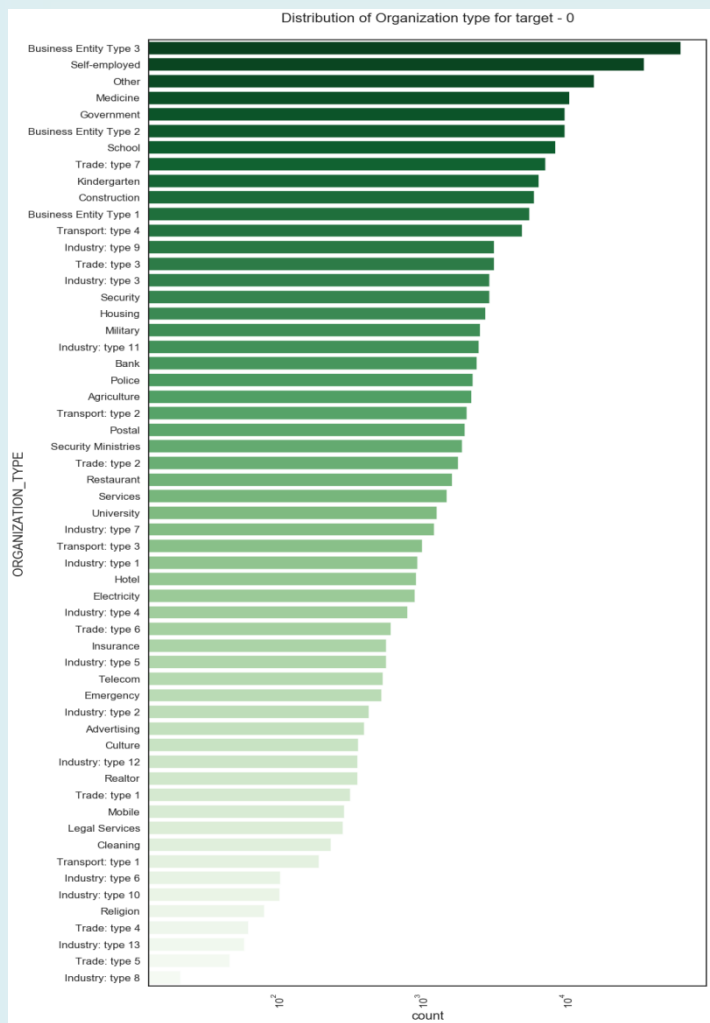


1. For income type 'working', 'commercial associate', and 'State Servant' has high income, slightly high in male.
2. For this Females are having more number of credits than male.
3. Less number of credits for income type 'Maternity leave'.
4. For Target 1 There is no income type for 'student', 'pensioner' and 'Businessman' which means they don't do any defaulters record



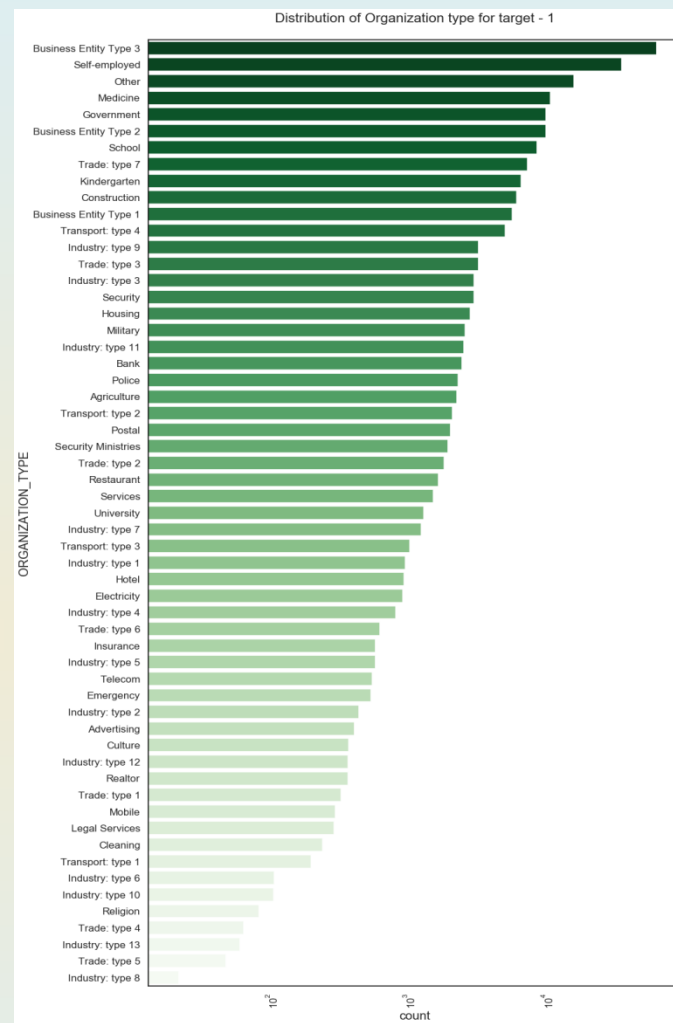
1. For contract type 'cash loans' is having higher number of credits than 'Revolving loans' contract type.
2. For this also Female is leading for applying credits.
3. There is only Females with Revolving loans in Target1.

Target0 (defaulters)



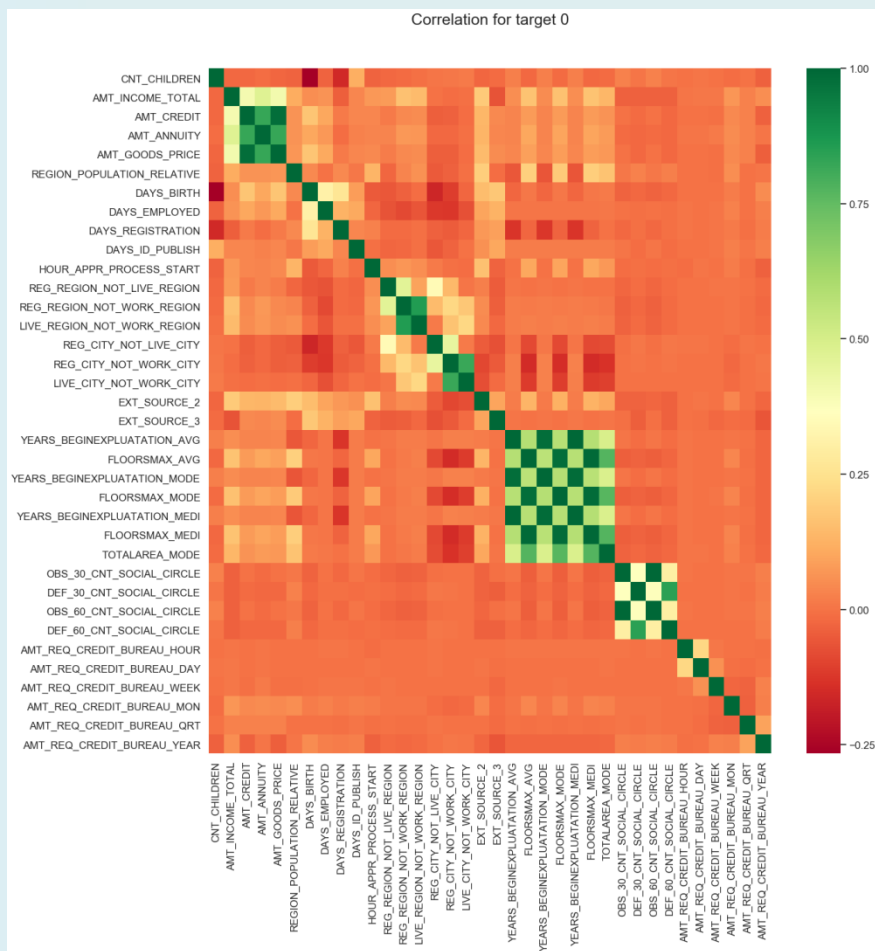
1. Employees from 'Business entity Type 3', 'Self employed', 'Other', 'Medicine', 'Government' and 'Business entity Type 2' are the top organizations types who applied for the credits.
2. Less employees are from Industry types, Religion, trade, cleaning, electrical Culture, hotel legal services, advertising organization type.

Target1 (non- defaulters)



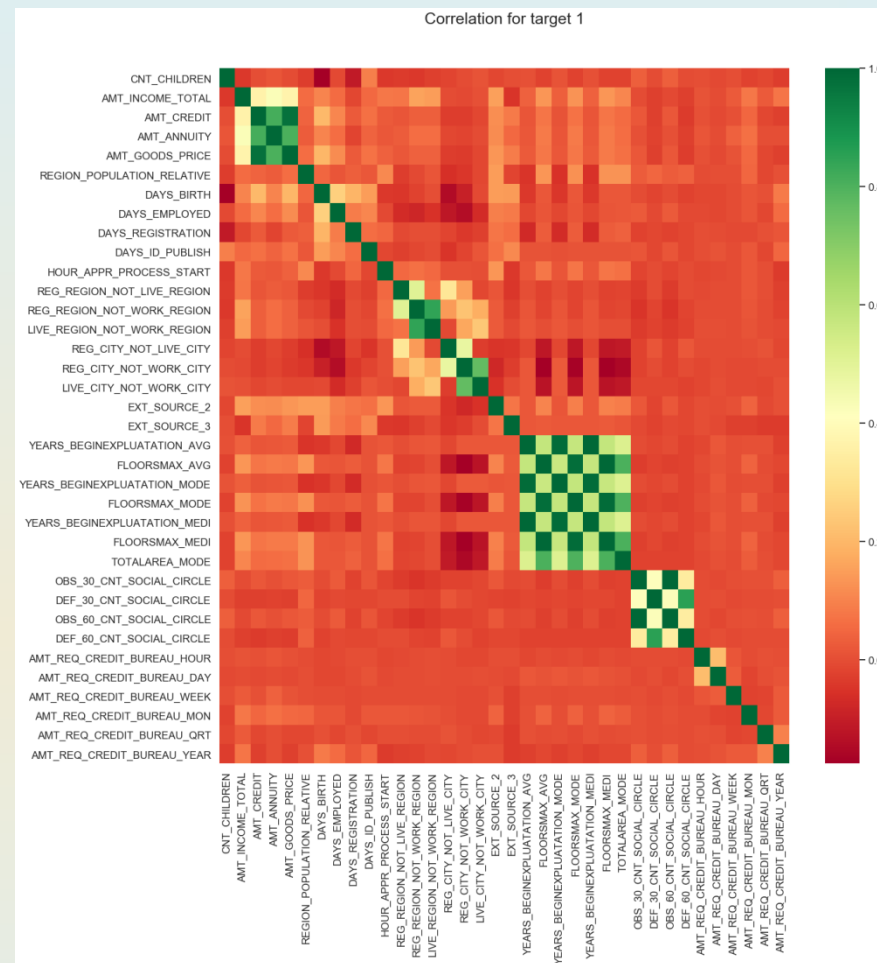
1. Employees from 'Business entity Type 3', 'Self employed', 'Other', 'Medicine', 'Government' and 'Business entity Type 2' are the top organizations types who applied for the credits.
2. Less employees are from Industry types, Religion, trade, cleaning, electrical Culture, hotel legal services, advertising organization type.
3. Same as Target0 in distribution of organization type.

Target0 (defaulters)



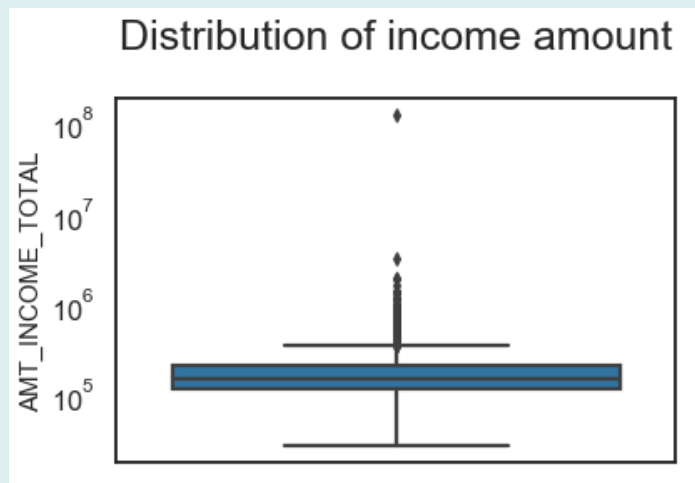
1. Amt_credit is inversely proportional to count of children, Days of birth, i.e. credit amount is higher for young age and less number of children or low dependents.
2. In target0 densely populated area has less count of children.

Target1 (non- defaulters)

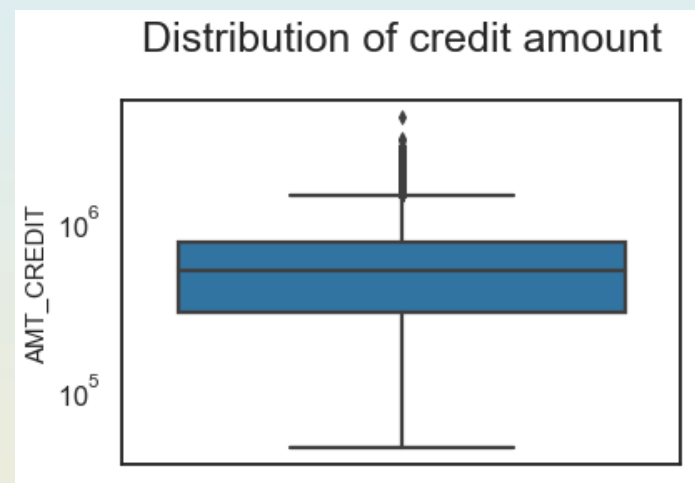


1. The heatmap is almost similar to Target0
2. The client's permanent address does not match work address

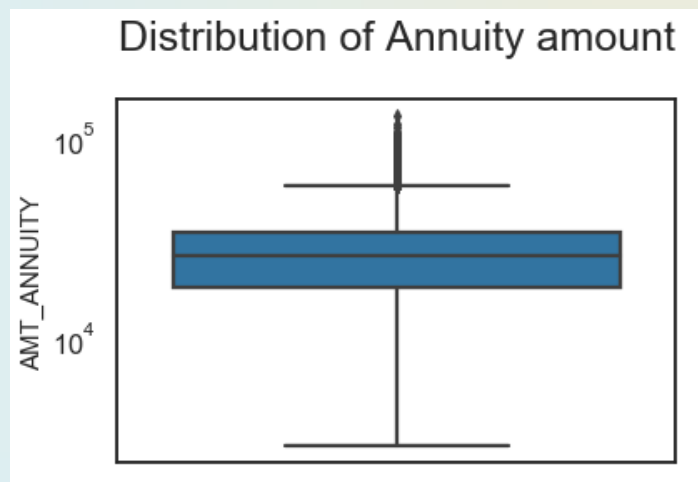
Target0 (defaulters)



1. few outliers are noticed in income amount.
2. The third quartiles is less for income amount.



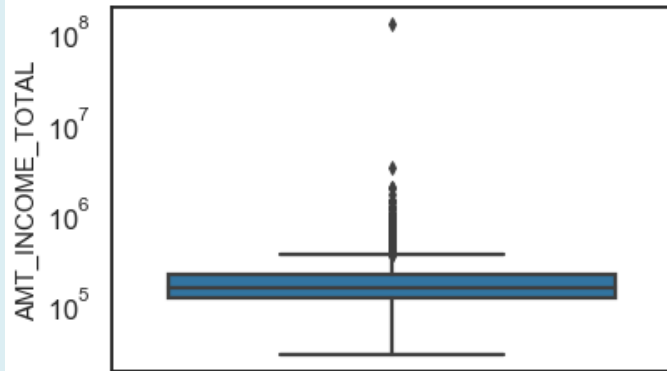
1. Some outliers are noticed in credit amount.
2. Most employees of target0 amount credit falls in first quartile since in credit amount first quartile is bigger.



1. Very few outliers are noticed in annuity amount.
2. Most employees of target0 amount credit falls in first quartile since in credit amount first quartile is bigger, same as credit amount employees graph.

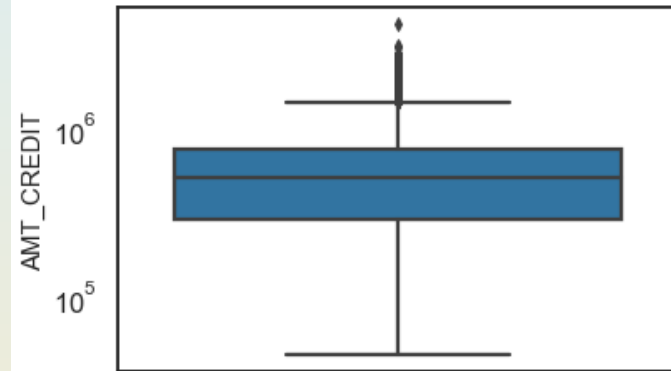
Target1 (non- defaulters)

Distribution of income amount



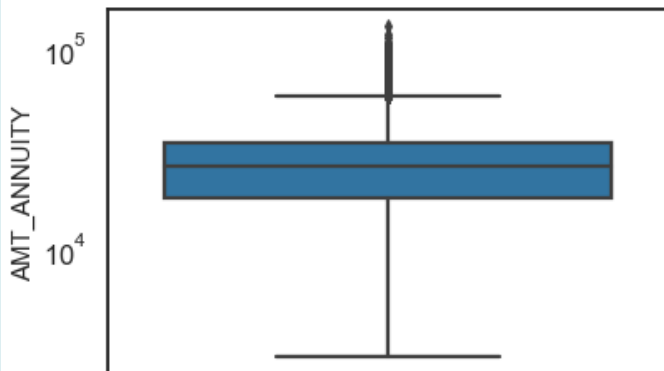
1. Very few outliers are noticed in income amount.
2. Most employees of target1 amount credit falls in first quartile since in credit amount first quartile is bigger

Distribution of credit amount



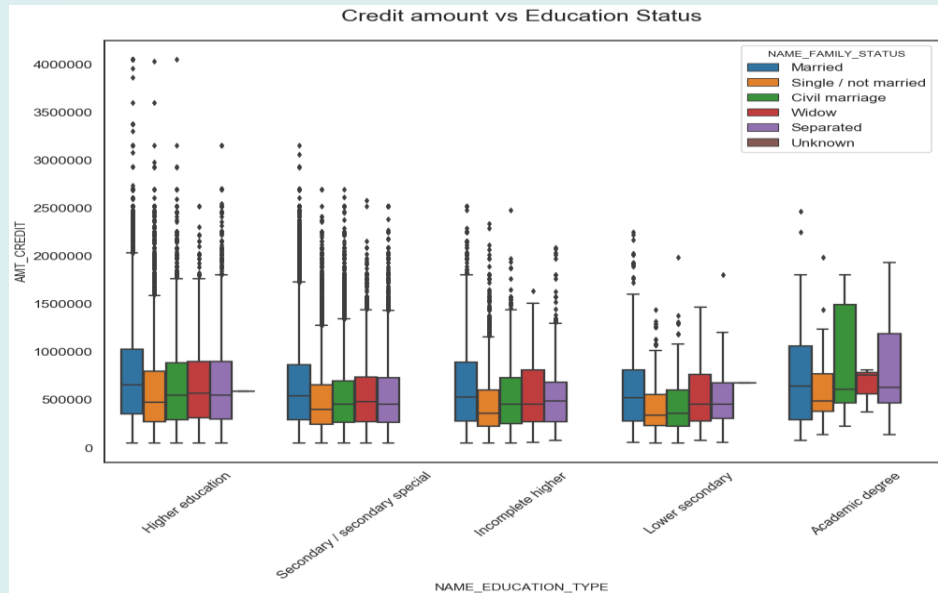
1. Very few outliers are noticed in amount credit.

Distribution of Annuity amount

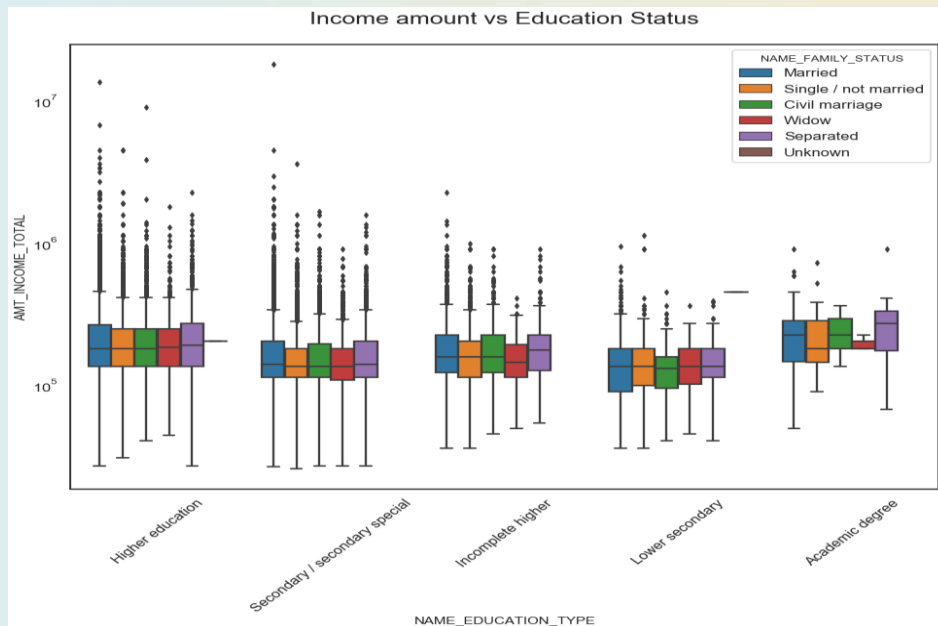


1. Very few outliers are noticed in annuity amount.
2. Most employees of target0 amount credit falls in first quartile since in credit amount first quartile is bigger, same as amount annuity employees in target0.

Target0 (defaulters)

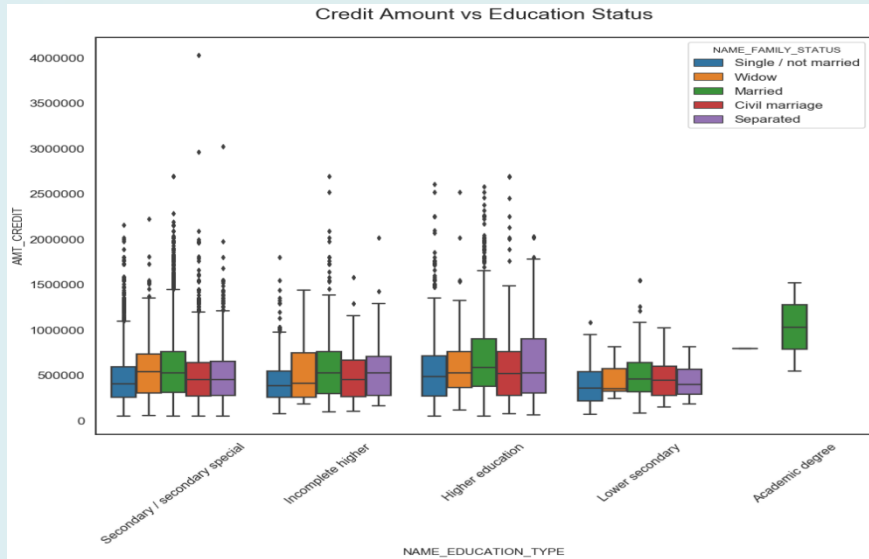


1. Civil marriage, Marriage and Separated employees of Academic degree are having higher number of credits.
2. very high outlier of Married, Single, and Civil marriage of Higher education is observed.

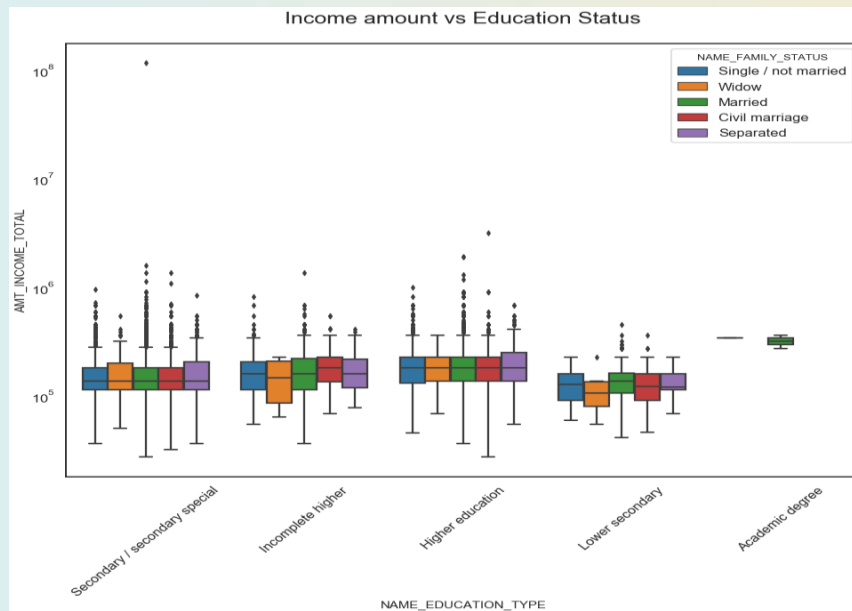


1. Median of Secondary, higher and lower secondary education is equal with family status in target0 dataset.
2. Most employees of higher and secondary education amount income total falls in first quartile since in total income amount first quartile is bigger
3. very high outlier of Married, Single, and Civil marriage of Higher education is observed.
4. In lower secondary education family status is having less income amount than others.

Target1 (non- defaulters)



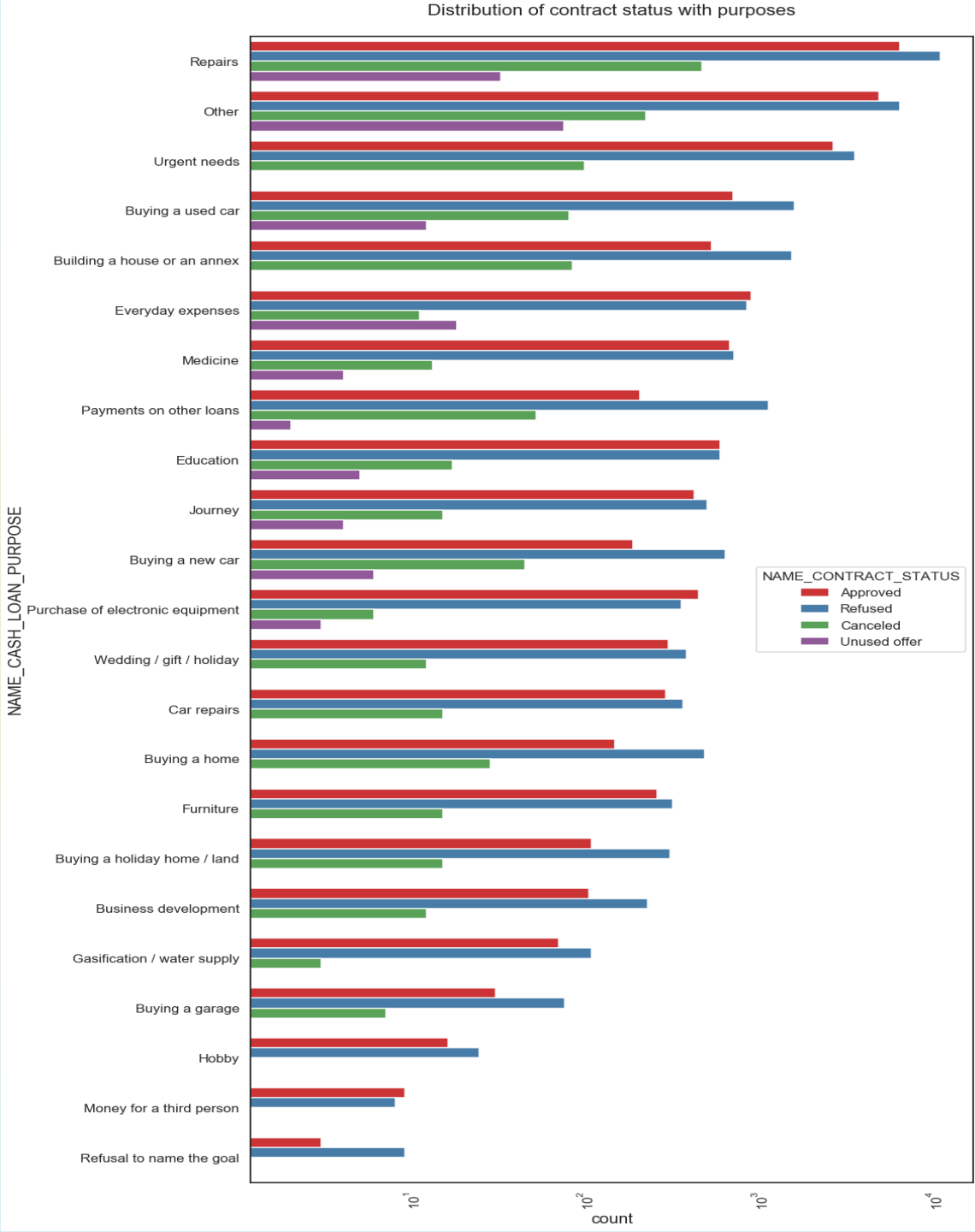
1. Civil marriage, Marriage and Separated employees of Academic degree are having higher number of credits.
2. Most of the outliers are from Education type 'Higher education' and 'Secondary' in all family status.



1. Median of Secondary, and higher education is equal with family status in target1 dataset.
2. Very few outliers present in target1.
3. The Academic degree has higher income in married employees.
4. In lower secondary education family status is having less income amount than others.

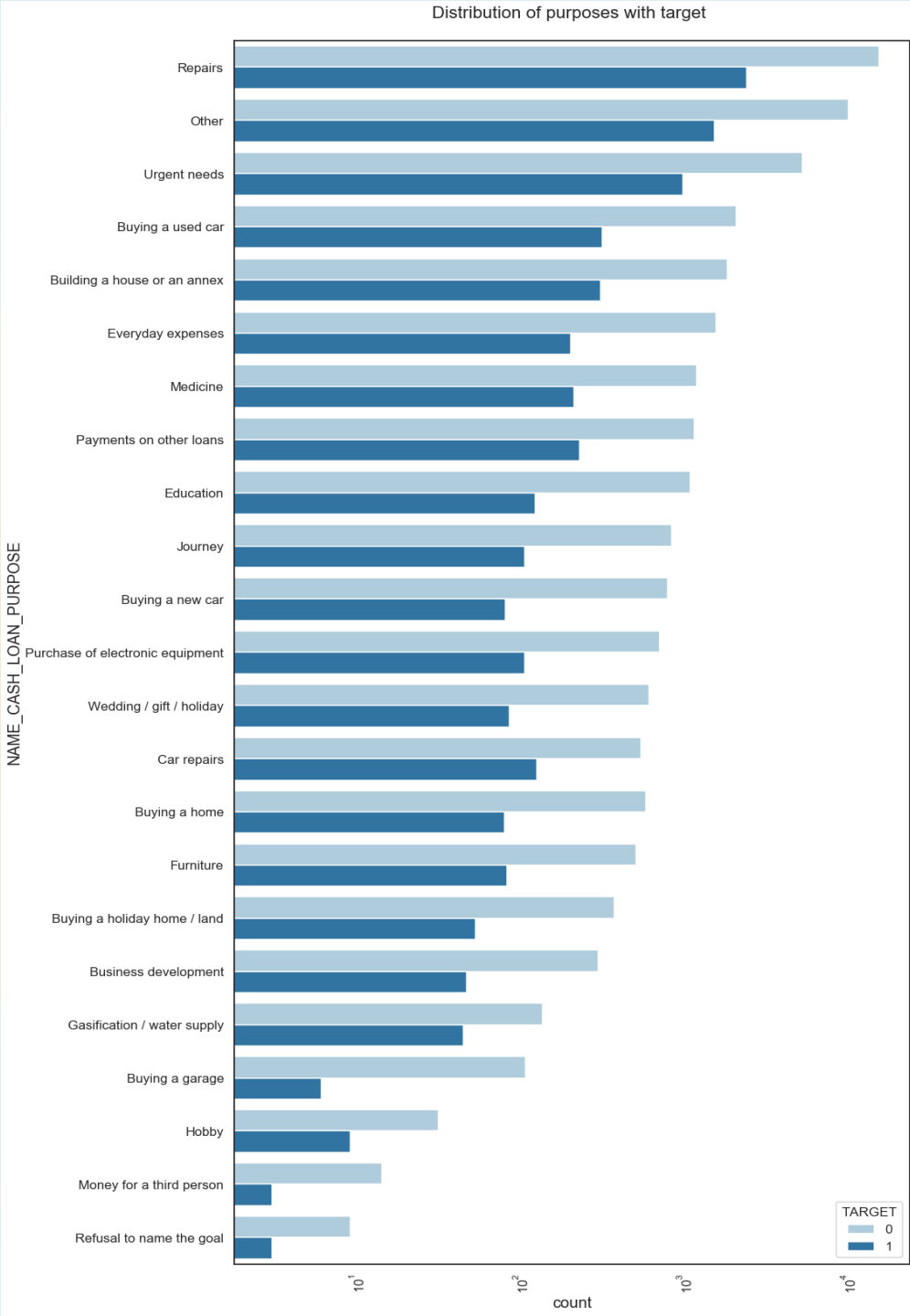
Distribution of contract status with purposes

- 1. Most rejection of loans came from purpose repairs followed by urgent needs, others, payments of other loans.
- 2. For education, Medicine and everyday expenses have, equal number of approvals and rejections.
- 3. Purchase of electronics equipment's have high approvals.



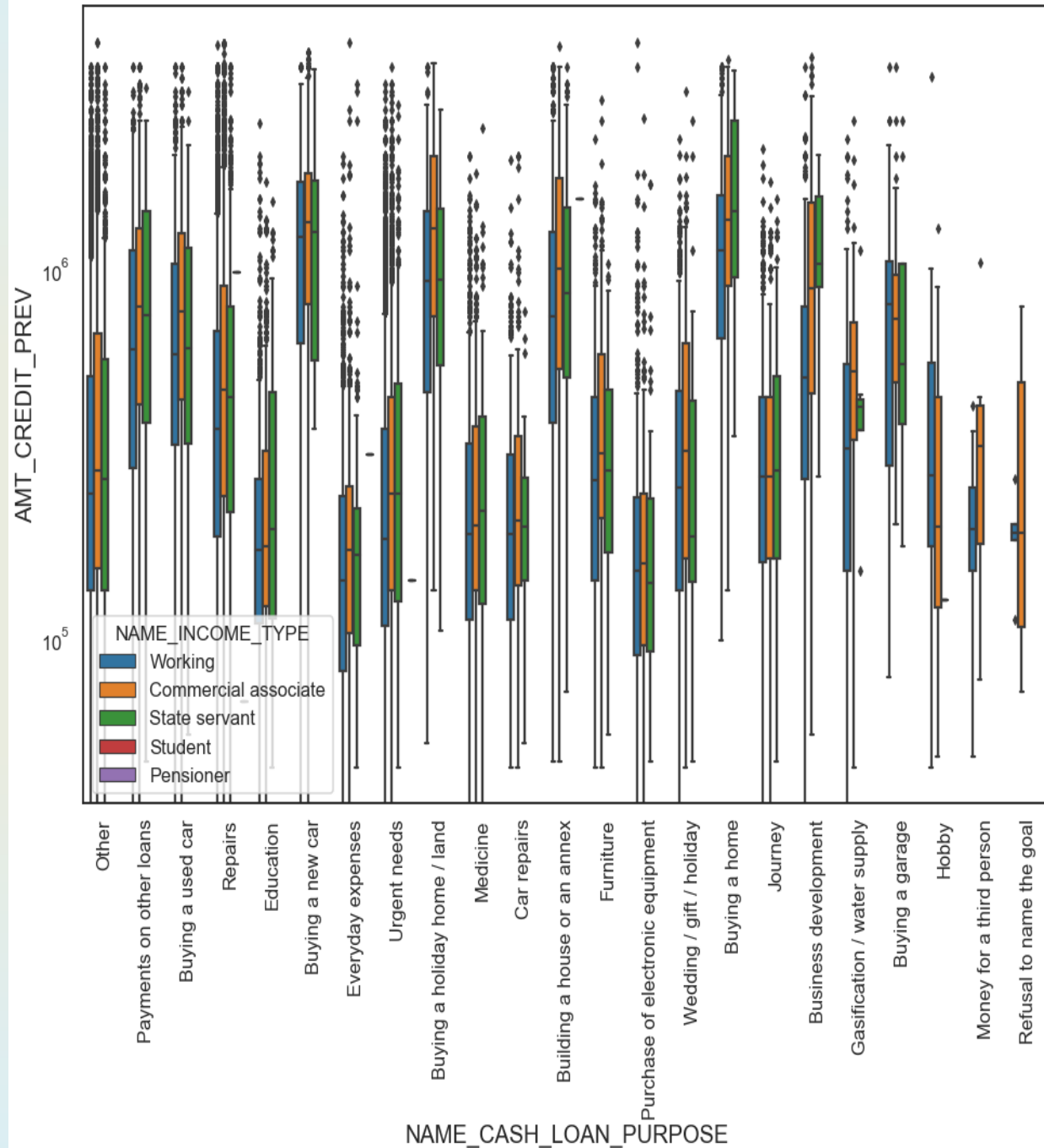
Distribution of purposes with Target 0 & Target 1

- 1. Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- 2. Buying a garage, business development, buying a land, buying a new car and Education are paying loans on time. Hence we can focus on these purposes for which the client is having for minimal payment difficulties.



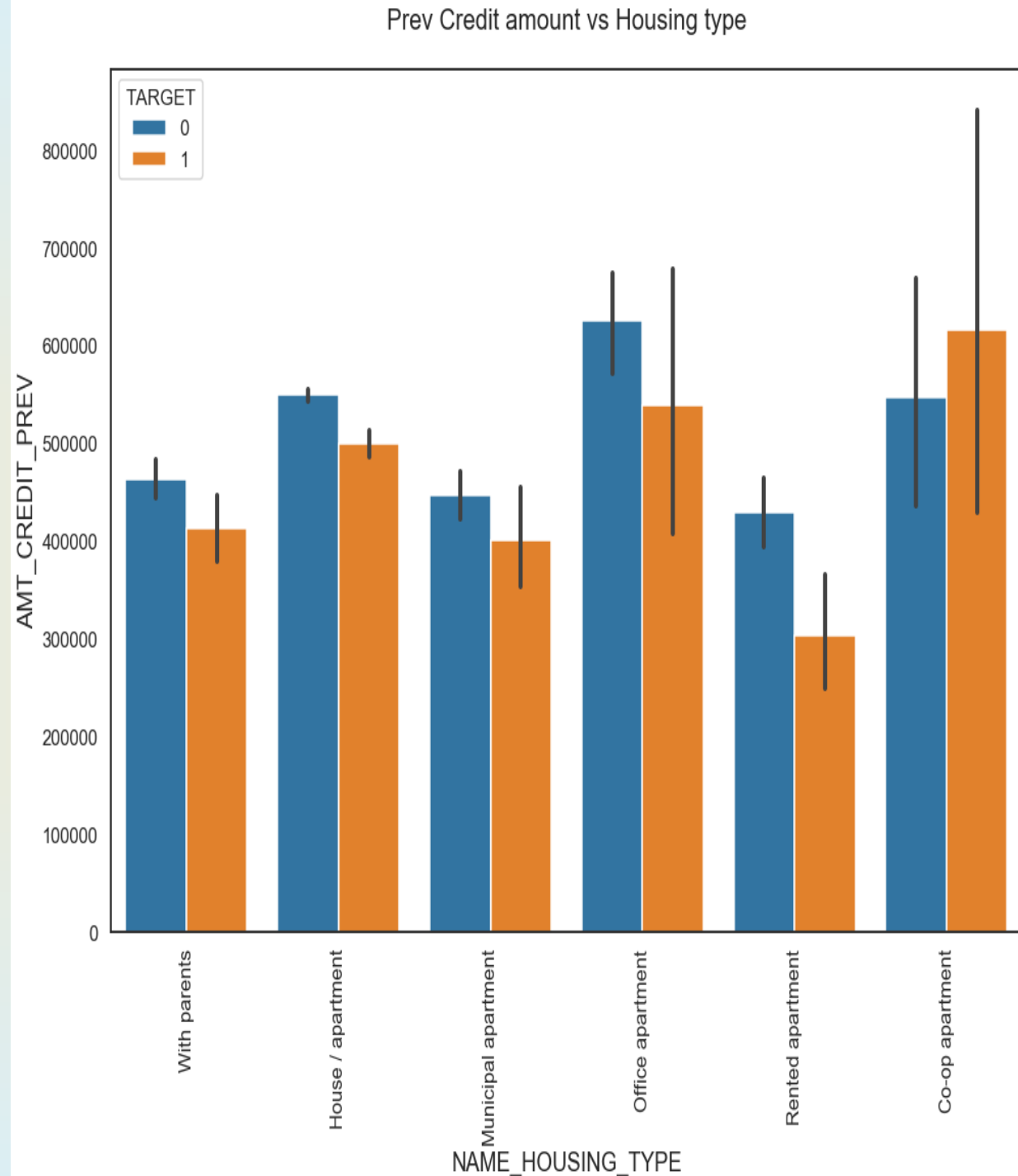
Previous credit amount Vs Loan Purpose

1. The credit amount of Loan purposes like 'Buying a home, Buying a land, Buying a new car and Building a house is higher.
2. Income type of state servants have a significant amount of credit applied.
3. Money for third person or a Hobby is having less credits applied for.



Previous credit amount Vs Housing type

1. Housing type, office apartment, Municipal apartment, and with parents is having higher credit of target 0
2. Bank should avoid giving loans to co-op apartment since Co-op apartment is having higher credit of target 1
3. Bank can focus mostly on the population with Rented housing type, since the defaulters population is less and more successful in paying the loans on time when compared with other housing type.



Conclusion

1. Banks should focus more on contract type 'Student' , 'pensioner' and 'Businessman' with housing 'type other than Co-op apartment' for successful payments.
2. Buying a garage, business development, buying a land, buying a new car and Education are paying loans on time. Hence we can focus on these purposes for which the client is having for minimal payment difficulties.
3. Banks should focus less on income type 'Working' as they are having most number of unsuccessful payments.
4. Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.
5. Get as much as employees from housing type 'With parents' as they are having high number of successful payments.
6. Bank can focus mostly on the population with Rented housing type, since the defaulters population is less and more successful in paying the loans on time when compared with other housing type.

Thank you